Working Title: A Comparison of Approaches for Unplanned Sample Sizes in Phase II Clinical Trials

By

Molly Olson

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biostatistics

May, 2017

Nashville, Tennessee

Approved (in progress):

Tatsuki Koyama , Ph.D.

Jeffrey Blume , Ph.D.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In this thesis, we develop .....

Chapter 1

Introduction

The introduction will talk about the motivation for the thesis. Introduce some examples of when we would need such methods.

Oncology phase II clinical trials are often used to evaluate the initial effect of a new regimen to determine if to warrant further study in a phase III clinical trial [1, 2, 3]. Simon's two-stage design [2] is a commonly used design in specifying sample sizes and critical values in phase II oncology clinical trials. Koyama and Chen [3] point out that it is common for actual sample sizes of these phase II trials to differ than the planned, pre-specified sample sizes. This could happen because of unanticipated accruement speed, drop-out rates are unexpected, and often multi-center trials can be slow in sharing information ... Currently, when attained sample sizes differ from planned, call these unplanned sample sizes, it is common practice to treat the attained sample sizes as planned. Though, when acheived sample sizes differ from planned, hypothesis testing using the attained sample sizes as if they were planned is not valid and hypothesis testing in these cases is not straightforward [1, 3]. Because of these reasons, extensions of Simon's design for hypothesis testing with unplanned sample sizes is important.

Talk about examples here

There have been many attempts to develop Frequentist methods that handle unplanned sample sizes in the second stage while using the planned stage I sample size, but my literature review found that there were only few Frequentist methods to handle unplanned sample sizes in both stage I and stage II. Likelihood based designs, that are able to be an extension of Simon's design, offer a nice solution to this problem because these designs offer flexibility in sample size without inflation of type I error. In this paper, we discuss the different methods for Simon's design when the attained stage II sample size is different than planned and when attained sample sizes in both stages are different than planned. In chapter 4, we review a concrete example from a Likelihood-based clinical trial, and in chapter 5, results of a numerical and theoretical study comparing the Frequentist properties of approaches in the setting where both stages differ **wording** in different settings are presented.

Chapter 2

Background

Simon's design will go here. This section will also talk about extending/shortening a trial (unplanned sample sizes) and how recalculating as if it were the planned design will introduce bias and inflate type I error. Talk about prespecifying (maybe here?)
We will only talk about extensions to Simon's design, hypothesis testing, and only stopping for futility in this paper.

Simon's II stage designs for clinical trials are common designs for phase II oncology clinical trials [2]. In Simon's designs, the null hypothesis $\mathrm{H}_0 : p \leq p_0$ is tested against the alternative $\mathrm{H}_1 : p \geq p_1$, where p is the true response probability, p_0 is the highest probability of response that would indicate that the research regimen is uninteresting and p_1 is the lowest probability of response that would indicate that the research regimen warrents further investigation. Under these hypotheses, it is required that the type I error rate remain less than α and power remain above $1-\beta$. The general framework of Simon's design includes a sample size and critical value in each of the two stages. Let n_1 denote the first stage sample size, n_t the sample size at the end of the second stage, r_1 the first stage critical value, and r_t the critical value for the end of the second stage. Let X_1 be the number of successes observed in the first stage and X_2 be the number of additional success in the second stage. In the first stage, n_1 patients are enrolled. If r_1 or fewer patients ($X_1 \leq r_1$) are successes, then the regimen is rejected and the trial is stopped for futility. If $r_1 + 1$ patients are successful, then the trial continues to the second stage. In the second stage, $n_2 = n_t - n_1$ patients are enrolled. If r_t or fewer out of the n_t patients are successful ($X_t = X_1 + X_2 \leq r_t$), the treatment is considered to be futile, otherwise if $r_t + 1$ patients succeed, the treatment is considered to be effective and warrent further study.

Design characteristics: Let B denote the cumulative binomial distribution function and b denote the binomial probability mass function. The probability of early termination with probability p in Simon's designs is given by PET = $B(r_1, p, n_1)$. The expected sample size for probability p is then EN $= n_1 + (1 - \text{PET})n_2$. The probability of rejecting a drug for probability p is then $PR(p) = B(r_1, p, n_1) + \sum_{x=r_1+1}^{min[n_1,r]} b(x, p, n_1)B(r-x, p, n_2)$. It is required that $PR(p) \geq 1 - \alpha$ and $PR(p) \leq \beta$. Given these constraints, it follows that unconditional conditional power, $UCP(p)$, given probability p, is given by $1 - PR(p) = 1 - \left(B(r_1, p, n_1) + \sum_{x=r_1+1}^{min[n_1,r]} b(x, p, n_1)B(r-x, p, n_2)\right) = \sum_{r_1+1}^{n_1} \left\{\sum_{x_2=r_t-x_1+1}^{n_2} b(x_2, p, n_2)\right\} b(x_1, p, n_1)$, and $UCP(p_1) \geq 1 - \beta$ and $UCP(p_0) \leq \alpha$.

Simon introduced optimal and minimax designs. An optimal two-stage design is a Simon's design in which minimizes the expected sample size under the null hypothesis, response value p_0, (EN_0) while still satisfying the type I and type II error probability restrictions. The minimax design will minimize the maximum sample size ($n_1 + n_2$). Jung *et al.* [4] introduced an extension of Simon's designs called admissible designs that are considered a compromise between optimal and minimax designs because they have similar maximim sample size as the minimax design and a similar EN_0 to the optimal design. Admissible designs optimize a straight line on the (n, EN)-plane, $q \times n + (1-q) \times \text{EN}$, for some $q \in [0, 1]$ [4]. Admissible designs satisfy (α, β) constraints and obtain an expected sample size somewhere between optimal and minimax designs. Admissible designs may be attractive because they have agreeable properties of both the minimax and optimal design. Simon does not allow for early termination of the trial for efficacy [2], and we do not consider that design here.

Chapter 3

Unplanned Sample Sizes

This chapter will talk about unplanned sample sizes when only the second stage is different and when both stages can be different. The former will talk about methods such as Koyama and Chen, UMVUE, MLE, etc. The latter will talk about the likelihood design, Chang, adaptation of Chang, and possibly Wu.

3.1 Unplanned Sample Sizes in Second Stage

3.2 Unplanned Sample Sizes in Both Stages

Decide later if subsections are needed.

3.2.1 *Chang et al.* Alternative Designs and Adaptation

Because accruement of patients can often be unexpected in both stages, it's imperative that methods are available to handle situations with attained sample sizes that differ from the planned sample size. Chang *et al* [5] proposed an alternative design that is an extension of Simons two stage design to handle unplanned sample sizes in both the first and second stages. Let n_1, n_t, r_1, r_t, p_0, p_1, α, and β be the original, planned design parameters **can we call these design parameters? Jung calls design parameters p0,p1,alpha,beta** and let n_1^* be the attained sample size in the first stage, n_t^* be the attained sample size at the end of the second stage, and $n_2^* = n_t^* - n_1^*$. Chang *et al* proposes that type II error probability spent in stage I, based on planned and attained sample size, is given by $\beta_1 = \mathrm{P}(X_1 \leq r_1 | n_1, p = p_1)$ Wu *et al* [6] also proposed an adjustment to Simons design based on attained sample sizes in both the first and second stages. Because Wu's methods don't work very well **wording**, we won't consider their method for the remainder of this paper.

3.2.2 Likelihood Design

Likelihood characteristics:

$$LR_n = \frac{L_n(\theta_1)}{L_n(\theta_0)} \in \{[0, 1/k], [1/k, k], [k, \infty)\}$$

Probability of Weak Evidence

$$\gamma_p = P(k_a \leq LR_n \leq K_b | H_p), k_a \leq k_b$$

Probability of Strong Evidence

$$\eta_1 = P(LR_n > k_b | H_1)$$

$$\eta_1 = P(LR_n < k_a | H_0)$$

Probability of Observing Misleading Evidence

$$\tau_0 = P(LR_n > k_b | H_0), \tau_0 \leq 1/k_b$$

$$\tau_1 = P(LR_n < k_a | H_1), \tau_1 \leq k_a$$

$\tau_i = O_p(n^{-1/2})$instead of remaining fixed like type I error

Translating likelihood properties into Simon-like design:

Interim: Translating to successes. This is the region in which we move to stage 2

$$UB_{interim} = \frac{log(k_{bi}) - n_1 log(\frac{1-p_1}{1-p_0})}{log(\frac{p_1(1-p_0)}{p_0(1-p_1)})}$$

$$LB_{interim} = \frac{log(k_{ai}) - n_1 log(\frac{1-p_1}{1-p_0})}{log(\frac{p_1(1-p_0)}{p_0(1-p_1)})}$$

(LB, UB) is the interval for weak evidence. If this was Simon's design, $LB_{interim} = r_1$

Probability of strong, misleading, and weak evidence under the null

$P(\text{Strong}_{0i}) = B(\lfloor LB_{interim} \rfloor, n_1, p_0)$

$P(Misleading_{0i}) = 1 - B(\lfloor UB_{interim} \rfloor, n_1, p_0)$

$P(Weak_{0i}) = B(\lfloor UB_{interim} \rfloor, n_1, p_0) - B(\lfloor LB_{interim} \rfloor, n_1, p_0)$

Probability of strong, misleading, and weak evidence under the alternative

$P(Strong_{0i}) = 1 - B(\lfloor UB_{interim} \rfloor, n_1, p_1)$

$P(Misleading_{0i}) = B(\lfloor LB_{interim} \rfloor, n_1, p_1)$

$P(Weak_{0i}) = B(\lfloor UB_{interim} \rfloor, n_1, p_1) - B(\lfloor LB_{interim} \rfloor, n_1, p_1)$

note: under Simon's, PET = 1-P(Weak)

Translating likelihood properties into Simon-like design:

Final Stage: Translating to successes.

The amount of successes that allow for continuation to the second stage are:

$$(\lfloor LB_{interim}+1\rfloor, \lfloor min(n_1, UB_{interim})\rfloor)$$

Probability of strong, misleading, and weak evidence under H_p

$$P(Weak_p) = \sum_{x=\lfloor LB_{interim}+1\rfloor}^{\lfloor min(n_1,UB_{interim})\rfloor} \Big(b(x,n_1,p_p)\times B(UB_{interim}-x,n-n_1,p_p)\Big) - B(LB_{interim}-x,n-n_1,p_p)$$

$$P(Strong_p) = P(Strong_{0i}) + \sum_{x=\lfloor LB_{interim}+1\rfloor}^{\lfloor min(n_1,UB_{interim})\rfloor} \Big(b(x,n_1,p_0)\times B(LB_{interim}-x,n-n_1,p_0)\Big)$$

$$P(Misleading_p) = P(Misleading_{0i}) + \sum_{x=\lfloor LB_{interim}+1\rfloor}^{\lfloor min(n_1,UB_{interim})\rfloor} \Big(b(x,n_1,p_p)\times(1-B(UB_{interim}-x,n-n_1,p_p)\Big)$$

If we want to translate likelihood design into a Simon's design, we overwrite the LR limits above as:

$$k_{ai} = OR^{r_1}\frac{1-p_1}{1-p_0}^{n_1} = \frac{1-p_0}{1-p_1}^{r_1-n_1}\frac{p_1}{p_0}^{r_1}$$

$$k_a = OR^{r}\frac{1-p_1}{1-p_0}^{n} = \frac{1-p_0}{1-p_1}^{r-n}\frac{p_1}{p_0}^{r}$$

$$k_{bi} = k_b = \infty$$

Chapter 4

Example

Here put the results of comparing the Chang et al paper and adaptation to the protocol of the study. We used Monte Carlo simulation to examine the performance of the study design of Chang et al...

Chapter 5

Results

The findings will be put here - primarily tables covering different combinations of Simon's designs and unplanned sample sizes.
Should maybe talk about how I couldn't replicate some results in the paper.
Some differences with the likelihood:
controlling type 1 error, but criteria is controlling PET - I think T1E will still be controlled.
Assuming stage II sample size and R2 is the same. We can add cohorts at the end of stage II. Talk about this.

- Chapter 3: design parameters
- when do we introduce notation?
- Particularly when describing other peoples' methods, how do you cite?
- If likelihood design can translate into simon's design and have better properties - can that be used as a tool to come up with these designs instead of using Chang's method? Or are they different because you're operating under frequentist vs likelihood inference?

[3]

REFERENCES

[1] R. Porcher and K. Desseaux, “What inference for two stage phase II trials?,” *BMC Medical Research Methodology*, vol. 12, p. 117, 2012.

[2] R. Simon, “Optimal two-stage designs for phase ii clinical trials,” *Controlled Clinical Trials*, vol. 10, pp. 1–10, 1989.

[3] T. Koyama and H. Chen, “Proper inference from simon’s two-stage designs,” *Statistics In Medicine*, vol. 27, pp. 3145–3154, 2008.

[4] S.-H. Jung, Y. Lee, K. Kim, and S. L. George, “Admissible two-stage designs for phase ii cancer clinical trials,” *Statistics In Medicine*, vol. 23, pp. 561–569, 2004.

[5]

[6]